

Vision-Language Models in Remote Sensing: Current progress and future trends

Year: 2023 | Citations: 139 | Authors: Congcong Wen, Yuan Hu, Xiang Li, Zhenghang Yuan, Xiao Xiang Zhu

Abstract

The remarkable achievements of ChatGPT and Generative Pre-trained Transformer 4 (GPT-4) have sparked a wave of interest and research in the field of large language models (LLMs) for artificial general intelligence (AGI). These models provide intelligent solutions that are closer to human thinking, enabling us to use general artificial intelligence (AI) to solve problems in various applications. However, in the field of remote sensing (RS), the scientific literature on the implementation of AGI remains relatively scant. Existing AI-related research in RS focuses primarily on visual-understanding tasks while neglecting the semantic understanding of the objects and their relationships. This is where vision-LMs (VLMs) excel as they enable reasoning about images and their associated textual descriptions, allowing for a deeper understanding of the underlying semantics. VLMs can go beyond visual recognition of RS images and can model semantic relationships as well as generate natural language descriptions of the image. This makes them better suited for tasks that require both visual and textual understanding, such as image captioning and visual question answering (VQA). This article provides a comprehensive review of the research on VLMs in RS, summarizing the latest progress, highlighting current challenges, and identifying potential research opportunities. Specifically, we review the application of VLMs in mainstream RS tasks, including image captioning, text-based image generation, text-based image retrieval (TBIR), VQA, scene classification, semantic segmentation, and object detection. For each task, we analyze representative works and discuss research progress. Finally, we summarize the limitations of existing works and provide possible directions for future development. This review aims to provide a comprehensive overview of the current research progress of VLMs in RS (see Figure 1), and to inspire further research in this exciting and promising field.